

Contenuti

- Architettura di Internet
- Principi di interconnessione e trasmissione

Tecnologie delle reti di calcolatori

- World Wide Web
- Posta elettronica
- Motori di ricerca
- *Netiquette*

Servizi Internet
(come funzionano e come usarli)

- Antivirus
- Personal firewall

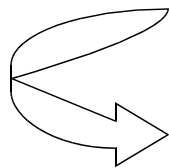
Servizi Internet
(come difendersi)

Applicazioni di rete e protocolli applicativi

- Tipicamente, ciascuna applicazione di rete (o servizio di rete) definisce un nuovo protocollo di livello applicativo ed un'interfaccia utente
- **Esempi**
 - **Trasferimento file**: usa protocollo applicativo **ftp**
 - **Collegamento a terminale remoto**: usa protocollo **telnet**
 - **World Wide Web**: usa protocollo **http**
 - **Posta elettronica**: usa protocollo **smtp**
 - **Chat**: usa protocollo **irc**
 - ...

Principali applicazioni Internet

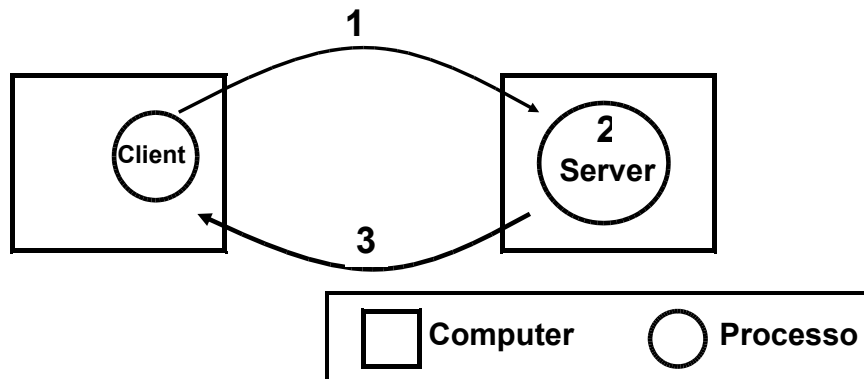
- **Domain Name System**
- **Posta elettronica (SMTP)**
- **Login remoto (Telnet)**
- **Trasferimento file (FTP)**
- **World Wide Web (HTTP)**



Tutte usano il modello *client/server*

Servizi Internet: *Modello client-server*

1. **Trasmissione di una richiesta dal processo client al processo server (*request*)**
2. **Elaborazione della richiesta da parte del server**
3. **Trasmissione di una risposta al client (*reply*)**



World Wide Web

WORLD WIDE WEB

La “killer application” di Internet

I chiari segnali del successo del Web

Yahoo, Netscape, Microsoft, Pointcast, AltaVista, CNN, ... *[Load misurato in hit]*
(>50 Milioni hits/day)

Evento	Periodo	Peak hits/day	Peak hits/minute
NCSA server (Oct. 1995)		2 Milioni	
Olympic Games 1996 (Atlanta, 1996)	180 Milioni	8 Milioni	
NASA Pathfinder (July 1997)	942 Milioni (14 days)	40 Milioni	
Olympic Winter Games (Japan, 1998)	634.7 Milioni (16 days)	55 Milioni	110.000
FIFA World Cup (France, 1998)	1.350 Milioni (90 days)	73 Milioni	209.000
Wimbledon (July, 1999)	942 Milioni (14 days)	125 Milioni	430.000
Wimbledon (July, 2000)		282 Milioni	964.000
Olympic Games 2000		875 Milioni	1.200.000

Elementi di Informatica, 2006/2007

7

I motivi alla base del successo del Web *("perché è risultata la killer application")*

- **Digitalizzazione dell'informazione**
 (Qualsiasi informazione codificata come sequenza di 0 e 1)
- **Diffusione di Internet** (dagli anni '70)
 (Trasporto dell'informazione ovunque, in tempi rapidissimi e a costi bassissimi)
- **Diffusione dei PC** (dagli anni '80)
 (Accesso, memorizzazione ed elaborazione dell'informazione da parte di chiunque a costi bassissimi)
- **Interfacce "user friendly"**

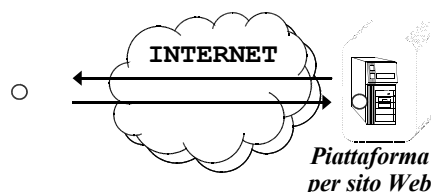
Elementi di Informatica, 2006/2007

8

Le 3 generazioni del Web (ieri)

- **1^a generazione = *Web publishing***

- Un ulteriore canale per informazione non critica
- 95% dell'informazione costituita da testo più eventuali immagini
- **Siti Web prevalentemente statici con alcune tecnologie (es., CGI) per funzionalità dinamiche**
- Manutenzione e aggiornamenti occasionali
- Canale economico
- Prestazioni molto variabili
- Affidabilità non garantita
- Sicurezza non indispensabile

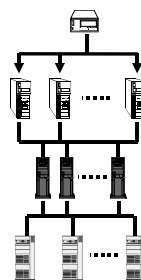


Le 3 generazioni del Web (ieri-oggi)

- **2^a generazione = *Web-based Information Systems***

- Canale di informazione critica, che sta diventando un mezzo di comunicazione privilegiato per molti utenti
 - “Vetrina” importante per industrie e organizzazioni
 - **Contenuti dinamici ed attivi in continuo aumento**
 - Interfaccia per molti altri servizi informatici usufruiti via rete (anche se non propriamente di rete)
 - ♦ Posta elettronica, trasferimento file, connessioni, ...
 - ♦ Accesso a archivi, basi di dati, banche dati, ...
 - Introduzione di servizi di streaming audio e video
 - Servizi personalizzati, servizi a pagamento, ...
- Quindi: necessità di qualità del servizio per
- ♦ **Prestazioni garantite (regola degli 8 secondi)**
 - ♦ **Affidabilità** (capacità di tollerare guasti)
 - ♦ **Sicurezza**

Piattaforma per sito Web

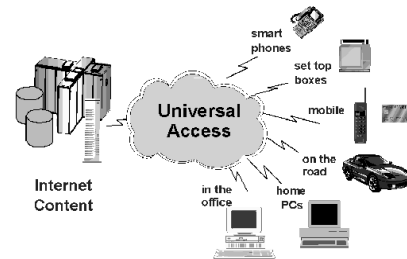


Le 3 generazioni del Web (oggi-domani)

- **3^a generazione = *Ubiquitous Web***

- Possibilità di usufruire di tutti i servizi Web della II generazione:

- ♦ *anytime* - sempre (24/7)
- ♦ *anywhere* - ovunque
- ♦ *anyway* - da qualunque dispositivo



Prima fase

- ♦ Si accettano prestazioni e affidabilità "variabili"
- ♦ **Sicurezza indispensabile**

Seconda fase

- ♦ Prestazioni e affidabilità garantita
- ♦ **Sicurezza indispensabile**

Ingredienti del Web

- **Meccanismi di comunicazione e naming propri di Internet**

- Stack di protocolli TCP/IP e Sistema DNS

- **Sistema client-server**

- **"Solo" tre nuovi standard**

- **URL**: Sistema di indirizzamento delle risorse
- **HTML**: Linguaggio di markup ipertestuale
- **HTTP**: Protocollo per le richieste risorse

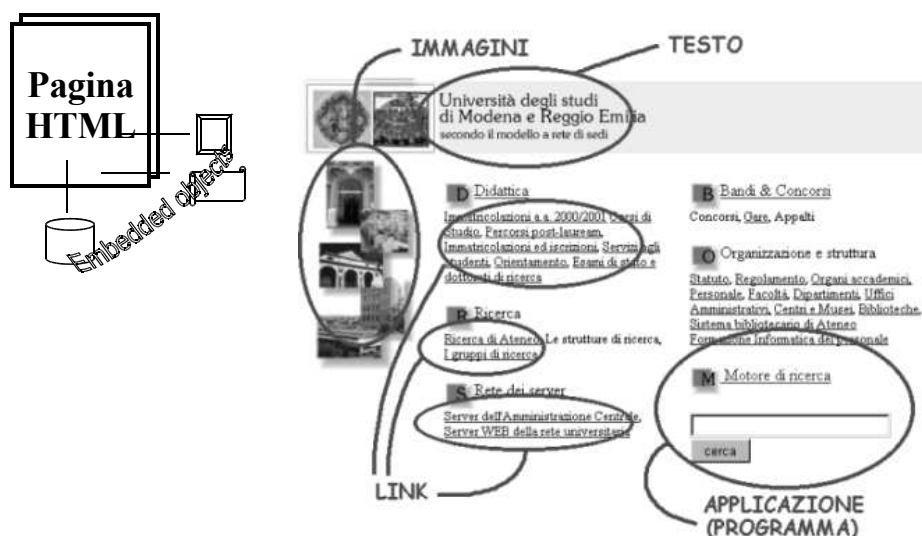
Uniform Resource Locator (URL)

Il sistema di indirizzamento delle risorse è basato su **Uniform Resource Locator (URL)**, un meccanismo standard per fare riferimento **a tutte le risorse** presenti nel Web:

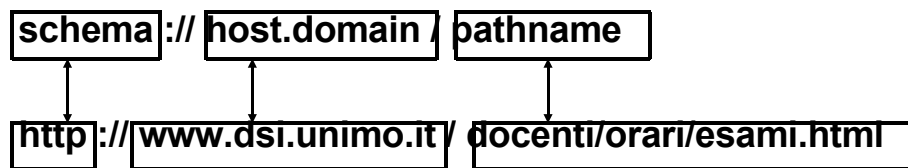
- **pagine** (testo, immagini, suoni, video, ...)
- **risultati di esecuzioni**
- **programmi eseguibili**

Tipica risorsa Web

- **Pagina HTML (contenitore) + embedded objects (file di qualunque tipo)**



Campi dell'URL



- **schema**: indica il modo con cui accedere alla risorsa, cioè quale protocollo bisogna usare per interagire con il server che controlla la risorsa. Il metodo di accesso più comune è **HTTP** (protocollo nativo del WWW per il recupero di risorse Web)
- **host.domain**: è l'hostname del nodo nel quale risiede la risorsa Web
- **pathname**: identifica la risorsa presso il server Web

Altre definizioni

- **Sessione utente**: serie di *richieste di risorse* effettuate dallo stesso utente al medesimo sito Web
- **Richiesta di risorsa (o pagina)**: una richiesta che tipicamente consiste di multipli *hit* inviati dal client dell'utente al sito Web
- **Hit**: una richiesta per un singolo oggetto effettuata dal client al server Web

Ingredienti del Web

- **Meccanismi di comunicazione e naming di Internet**
 - Protocollo TCP/IP e Sistema DNS
- **Sistema client-server**
- **Informazione digitalizzata**
- **“Solo” tre nuovi standard**
 - **URL**: Sistema di indirizzamento delle risorse
 - **HTML**: Linguaggio di markup ipertestuale
 - **HTTP**: Protocollo per le richieste risorse

Documento Web

- **Una pagina è costituita da vari oggetti (testo, immagini binarie, ...), detti *embedded objects*.**
- **Ad ogni oggetto corrisponde un file.**
- **Caratteristiche del “testo tipico”:**
 - Rappresentazione in standard ASCII
 - Si specifica sia il contenuto sia la rappresentazione (layout)
 - Scritto nel **linguaggio di markup HTML**

Concetto di Ipermedia

- **I documenti Web tipicamente contengono un insieme di**
 - testo
 - immagini
 - puntatori selezionabili ad altre risorse
 - (audio)
 - (video)

Ipermedia



- **Uso mediante “point-and-click”**

Linguaggio di markup

Per permettere la corretta visualizzazione dell'informazione su qualsiasi piattaforma hardware connessa in rete è stato necessario definire un nuovo linguaggio di markup per la formattazione delle pagine:

- Fornisce delle linee guida generali per la rappresentazione del contenuto.
- Non specifica esattamente il formato e la posizione del testo, lasciando ai browser la definizione dei dettagli.

Due browser potrebbero visualizzare lo stesso documento in modo differente

Linguaggio HTML

- Sebbene siano stati proposti modifiche ed altri standard*, a tutt'oggi *HyperText Markup Language* (HTML) rimane il “**linguaggio del Web**”
- La pagina HTML è un file di solo testo ASCII
- Il testo è *free-format*
- Contenuto del testo e specifiche di formato sono inseriti nello stesso file

* **XML è il nuovo standard più importante**

Istruzioni HTML

- La descrizione dei contenuti dell'ipertesto viene effettuata inserendo all'interno del testo stesso alcune istruzioni dette *marcatori* o *markup* o *tag* che producono le visualizzazioni e le azioni specificate.
- Le istruzioni HTML sono racchiuse tra parentesi angolari, nella forma **<tag>**, e vengono terminate da un tag di chiusura nella forma **</tag>**. Es.

```
<CENTER>  <B>  <FONT="Arial">
<FONT COLOR="#336633"> <FONT SIZE=+1>
Informazioni generali
</FONT> </FONT> </FONT> </B> </CENTER>
```

L'importanza del *tag àncora*

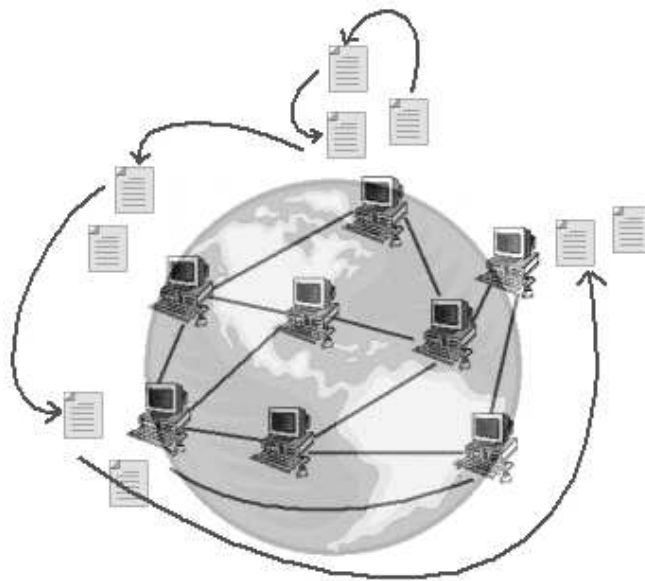
- L'istruzione più innovativa dell'HTML è l'**àncora** delimitata dai tag **<A>...**, in quanto tale elemento permette di trasformare un normale testo in ipertesto multimediale.
- Un àncora può far riferimento ad una sezione della stessa pagina oppure **ad una qualsiasi risorsa** (testuale, multimediale, eseguibile) presente sul Web, denotata mediante un **URL** che va inserito all'interno del tag àncora.

<A

```
  HREF="http://www.unimo.it/studenti/erasmus.ht  
  ml"> Programma Erasmus </A>
```

- Il testo **Programma Erasmus** viene visualizzato in modo differente e risulta un link simbolico selezionabile via mouse.

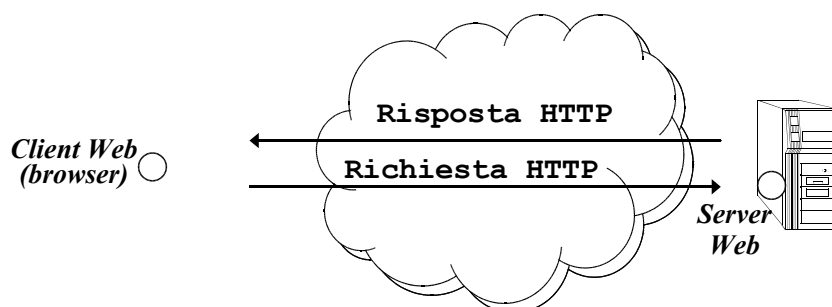
Il tag àncora consente l'ipermedialità su scala geografica



Ingredienti del Web

- **Meccanismi di comunicazione e naming di Internet**
 - Protocollo TCP/IP e Sistema DNS
- **Sistema client-server**
- **Informazione digitalizzata**
- **“Solo” tre nuovi standard**
 - **URL**: Sistema di indirizzamento delle risorse
 - **HTML**: Linguaggio di markup ipertestuale
 - **HTTP**: Protocollo per le richieste di risorse

Richiesta per una risorsa Web



Protocollo HTTP

- ***HyperText Trasmission Protocol* (HTTP)** è il protocollo che permette il reperimento delle risorse Web
- E' un protocollo applicativo di tipo **request/reply** basato sulla suite di protocolli TCP/IP
- Tutti i client Web e i server Web devono supportare il protocollo HTTP per poter scambiare richieste e risposte. Per questa ragione i **client** e i **server Web** sono chiamati anche **client HTTP** e **server HTTP**

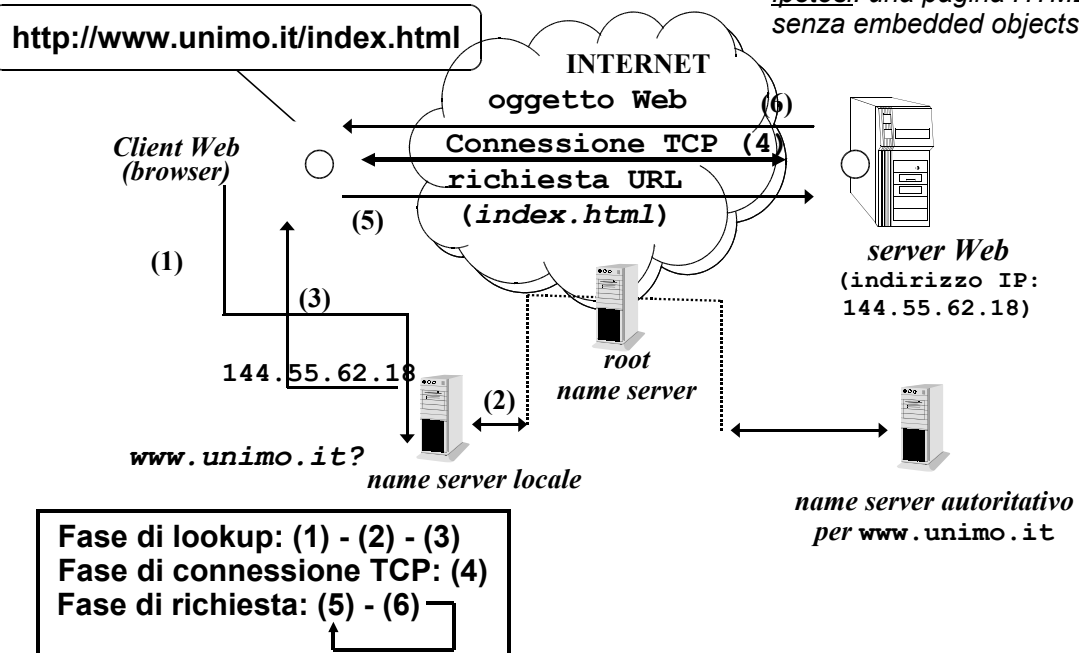
Protocollo HTTP

- Il browser richiede l'apertura di una connessione TCP verso il server Web
- Se possibile, il server Web accetta la connessione TCP dal client
- Una volta aperta la connessione, vengono scambiati vari messaggi HTTP di tipo testuale tra browser e Web server
- Al termine, vi è la chiusura della connessione TCP

Richiesta di una risorsa Web

(cosa succede ad ogni click del mouse?)

*Ipotesi: una pagina HTML
senza embedded objects*



Elementi di Informatica, 2006/2007

29

Approfondimenti

Principali componenti del Web

- ***Client Web (browser) o client HTTP***
- ***Server Web o server HTTP***
- **Protocollo HTTP**

Browser

- **Il browser è un'applicazione software che svolge il ruolo di interfaccia fra l'utente ed il WWW, mascherando la complessità di Internet.**
- **Diventa un client Web per recuperare informazioni dai server Web.**
- **Funzioni principali**
 - instaura una connessione TCP con il server tramite cui invia opportuni messaggi al server Web per ottenere le risorse richieste
 - interpreta il codice ipertestuale HTML
 - elabora il codice allo scopo di visualizzare in modo appropriato il contenuto delle pagine sullo schermo

Breve storia

- **FrameMaker e Acrobat PDF**
 - consentono generazione e incorporazione di iperlink.
- **Gopher** (University of Minnesota)
 - primo browser con iperlink verso siti remoti
 - uso di standard, quali testo ASCII e socket Unix
 - difetti: link separati dal testo, immagini non gestite

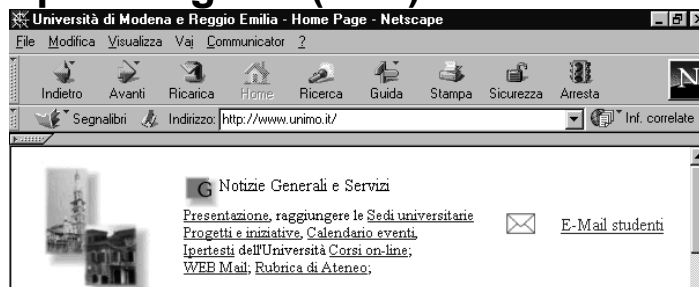
HTML

Mosaic

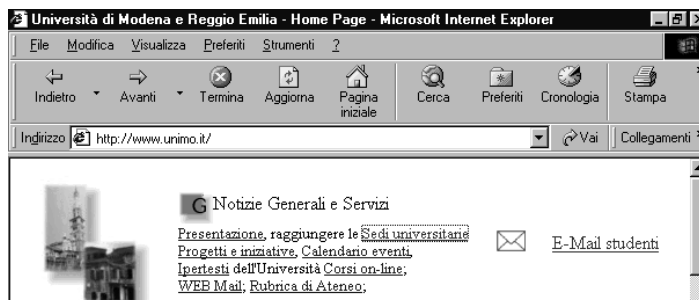
- **Mosaic** (NCSA, 1993)
 - ↳ **Netscape**
 - ↳ **Microsoft Explorer** (tramite *Spyglass Co.*)

Browser più diffusi

- **Netscape Navigator (AoL)**















- **Internet Explorer (Microsoft)**



Finestra del browser

- **Barra del titolo della pagina scaricata**
- **Barra dei menu delle funzioni del browser suddivise in classi**
 - File, Edit, View, Go, Help
- **Barra degli strumenti veloci**
 - per funzioni frequenti: back, reload, home, ...
- **Finestra indirizzo URL**
- **Area di visualizzazione pagina Web**
- **Barra di stato**
 - per messaggi all'utente: URL dell'iperlink, informazioni su operazioni in corso, ...

Pulsanti azione dei browser

Netscape	Explorer	Azione
 Indietro	 Indietro	Ritorna all'ultima pagina Web visitata precedentemente
 Arresta	 Termina	Interrompe il trasferimento della pagina Web corrente
 Ricarica	 Aggiorna	Richiede nuovamente al server Web la risorsa visualizzata, per aggiornare eventuali modifiche della pagina
 Home	 Pagina iniziale	Richiama la home page per il browser, o pagina di partenza, scelta dall'utente
 Ricerca	 Cerca	Attiva un motore di ricerca per trovare risorse che contengono specifiche parole
 Segnalibri	 Preferiti	Consente di memorizzare una serie di indirizzi di pagine Web scelte dall'utente

Funzioni di un browser

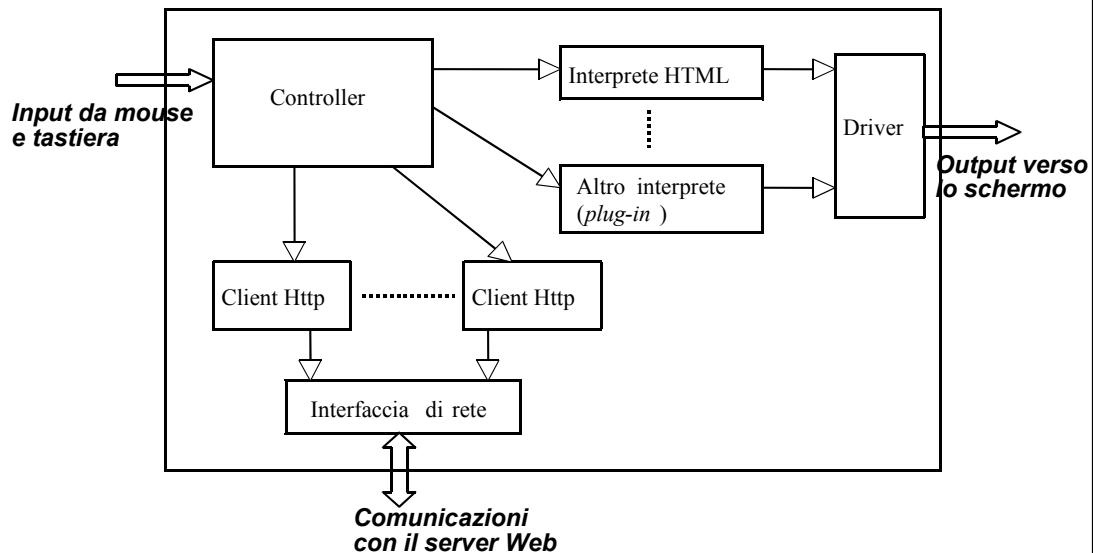
- Richiedere la risorsa al server Web
- Ricevere il file con la codifica della risorsa in linguaggio HTML
- Decodificare ed interpretare secondo le specifiche di HTML le caratteristiche grafiche, di formato e di comportamento dei vari oggetti contenuti nella risorsa (titoli, testi, immagini, pulsanti ...)
- Visualizzare la risorsa sullo schermo del computer dell'utente

Richiesta di una risorsa

- Una qualsiasi risorsa Web può essere richiesta o attraverso un link presente su un'altra risorsa, oppure digitando direttamente il suo indirizzo nella barra dell'indirizzo del browser.



Componenti interne di un browser



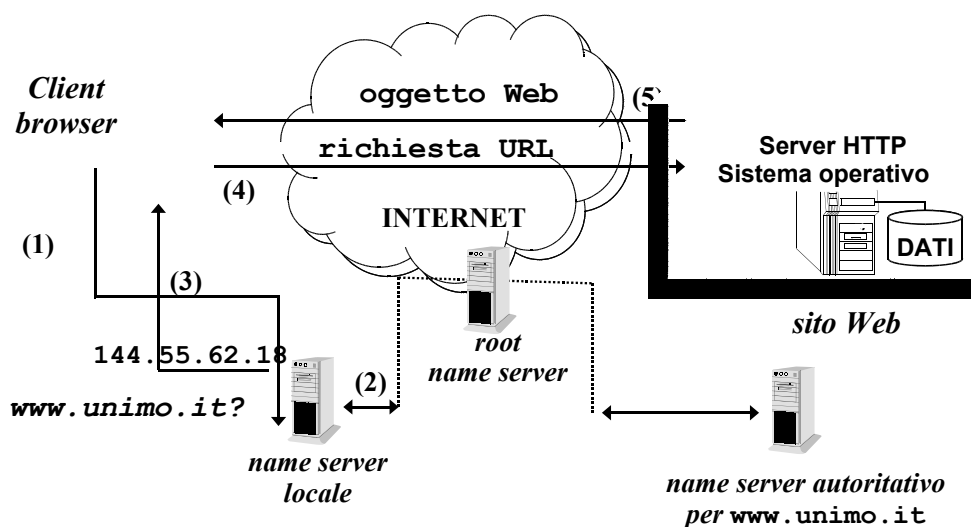
Principali componenti del Web

- **Client Web (browser) o client HTTP**
- **Server Web o server HTTP**
- **Protocollo HTTP**

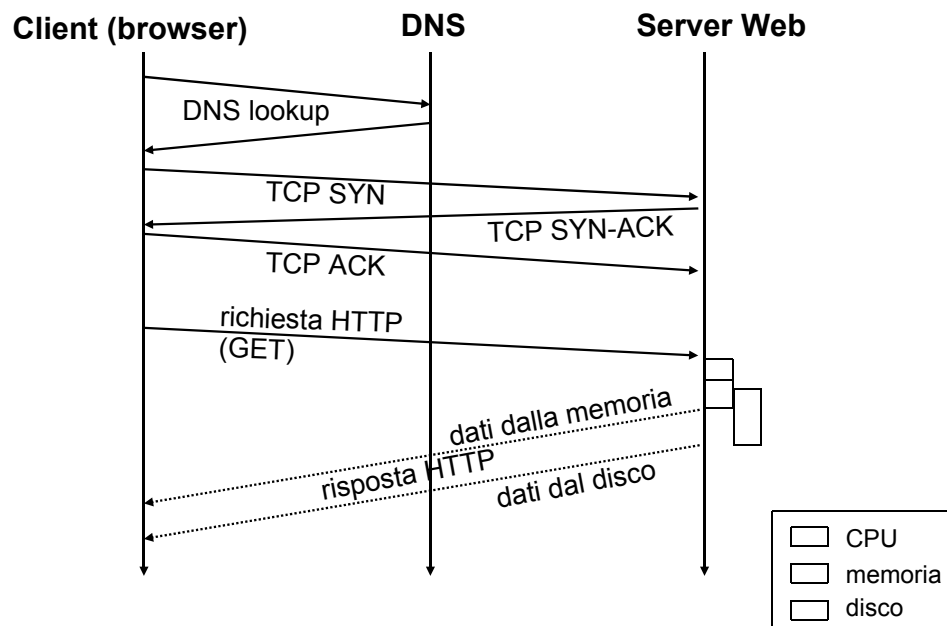
Componenti di un sito Web

- **Piattaforma hardware**
- **Software di base (*sistema operativo*)**
- **Server HTTP**
 - Il processo server HTTP ed il relativo software che viene eseguito sulla piattaforma stabilisce il collegamento tra la piattaforma (hardware - software di base) e la parte informativa del sito Web
- **Parte informativa e servizi del sito Web**
 - Il sito deve mettere a disposizione un insieme di risorse Web che possono essere richieste dai client con cui vengono instaurate delle connessioni HTTP

Architettura del sito Web “statico”



Server HTTP (es., richiesta oggetto statico)



Gestione richieste multiple

- Il server deve specificare una porta che identifica il servizio sull'host
- Tuttavia, più client possono richiedere il servizio in rapida successione
- Due soluzioni:
 - **Accodamento della richiesta client arrivata dopo**
Gestito automaticamente dal sistema operativo; il processo server deve specificare solo la lunghezza (*backlog*) della coda
 - **Gestione contemporanea di più richieste client**
Possibile mediante la gestione del multitasking da parte del sistema operativo

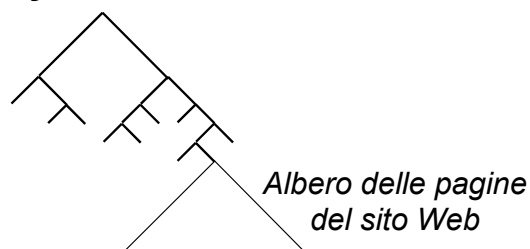
Componenti di un sito Web

- **Piattaforma hardware**
- **Software di base (*sistema operativo*)**
- **Server HTTP**
 - Il processo server HTTP ed il relativo software che viene eseguito sulla piattaforma stabilisce il collegamento tra la piattaforma (hardware - software di base) e la parte informativa del sito Web.
- **Parte informativa e servizi del sito Web**
 - Il sito deve mettere a disposizione un insieme di risorse Web che possono essere richieste dai client con cui vengono instaurate delle connessioni HTTP.

Organizzazione delle pagine

- Ogni pagina Web ha un nome unico, che corrisponde al cammino assoluto dalla radice “/” dell’albero delle pagine. (Questo cammino è proprio quello specificato nella parte *pathname* dell’URL richiesto dal client.)

File system del server



Tipi di risorse - 1 **(classificazione sulla base del contenuto)**

- pagina HTML
- testo in formato ASCII
- pagina preformattata (come PostScript, PDF)
- immagine in diversi formati (tipo GIF, JPEG)
- suono codificati in diversi formati (quali AU, AIFF, MP3)
- video in diverse rappresentazioni (quali Quicktime, MPEG)
- rappresentazione VRML di scene tridimensionali
- codice eseguibile in linguaggi interpretati (tipo Perl e shell)
- codice eseguibile in linguaggi compilati, tipo C
- codice Java

Tipi di risorse - 2 **(classificazione funzionale)**

Risorse statiche

pagine il cui contenuto è relativamente stabile nel tempo.

Risorse volatili

pagine il cui contenuto viene modificato da eventi in corso.

Es., ultime notizie, avvenimenti sportivi, titoli in borsa.

Risorse dinamiche

risorse il cui contenuto è creato dinamicamente sulla base della richiesta del client.